第一章

条件期望、条件方差和线性投影

Outline

1. 条件期望
2. 条件方差
3. 线性投影

条件期望的定义

令 x 为随机向量，y 为一元随机变量。以 x 为条件，y 的条件均值函数 (Conditional Expectation Function, CEF) 定义如下：

$$\mu(\mathbf{x}) = \mathbb{E}(\mathrm{y}|\mathbf{x}) = \begin{cases} \int_{\mathrm{y}} y \cdot f(y|\mathbf{x}) dy, & \mathrm{y} \text{ 连续}, \\ \sum_{\mathrm{y}} y \cdot P_{\mathrm{y}|\mathbf{x}}(y|\mathbf{x}), & \mathrm{y} \text{ 离散}, \end{cases}$$

其中，$f(y|\mathbf{x})$ 是条件概率密度函数，$P_{\mathrm{y}|\mathbf{x}}(y|\mathbf{x})$ 是条件概率。

条件期望的性质

迭代期望律（Law of Iterated Expectations, LIE）

$$\mathbb{E}_{\mathbf{x}}\left[\mathbb{E}[\mathrm{y}|\mathbf{x}]\right] = \mathbb{E}[\mathrm{y}]$$

Proof.

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}}\left[\mathbb{E}[\mathrm{y}|\mathbf{x}]\right] &= \int_{\mathbf{x}}\left(\int_{\mathrm{y}} y \cdot f(y|x)dy\right) f(x)dx \\
&= \int_{\mathbf{x}}\int_{\mathrm{y}} yf(x,y)dxdy \\
&= \int_{\mathrm{y}} yf(y)dy = \mathbb{E}[\mathrm{y}].
\end{aligned}$$

□

一般形式的迭代期望律

令 $\mathbf{w}$ 为随机向量，y 是一元随机变量，$\mathbf{x}$ 是随机向量且满足 $\mathbf{x} = \mathbf{f}(\mathbf{w})$.[a] 一般形式的 LIE 可表示为：

$$\mathbb{E}(y|\mathbf{x}) = \mathbb{E}\left[\mathbb{E}(y|\mathbf{w})|\mathbf{x}\right].^{b}$$

"The smaller information set always dominates."

特例：$\mathbb{E}(y|\mathbf{x}) = \mathbb{E}\left[\mathbb{E}(y|\mathbf{x}, \mathbf{z})|\mathbf{x}\right]$.

[a] $\mathbf{x} = \mathbf{f}(\mathbf{w})$ implies that if we know the outcome of $\mathbf{w}$, then we know the outcome of $\mathbf{x}$.

[b] Note that we also have $\mathbb{E}(y|\mathbf{x}) = \mathbb{E}\left[\mathbb{E}(y|\mathbf{x})|\mathbf{w}\right]$.

条件期望的线性性质

$$\mathbb{E}\left(\sum_{j=1}^{G} a_j(\mathbf{x})y_j + b(\mathbf{x})|\mathbf{x}\right) = \sum_{j=1}^{G} a_j(\mathbf{x})\mathbb{E}(y_j|\mathbf{x}) + b(\mathbf{x}).$$

The CEF Decomposition Property

For any random variable y, we have

$$\mathrm{y} = \mathbb{E}(\mathrm{y}|\mathbf{x}) + \varepsilon,$$

where

- ε is mean independent of $\mathbf{x}$, that is, $\mathbb{E}(\varepsilon|\mathbf{x}) = 0$,
- ε is uncorrelated with any function of $\mathbf{x}$.

Proof.

First, note that

$$\mathbb{E}(\varepsilon|\mathbf{x}) = \mathbb{E}[\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x})|\mathbf{x}] = \mathbb{E}(\mathrm{y}|\mathbf{x}) - \mathbb{E}(\mathrm{y}|\mathbf{x}) = 0.$$

Next, by property LIE, for any function $g(\cdot)$,

$$\mathbb{E}[g(\mathbf{x})\varepsilon] = \mathbb{E}(\mathbb{E}[g(\mathbf{x})\varepsilon|\mathbf{x}]) = \mathbb{E}(g(\mathbf{x})\mathbb{E}[\varepsilon|\mathbf{x}]) = 0.$$

□

The CEF Prediction Property

If $\mathbb{E}(\mathrm{y}^2) < \infty$ and $\mu(\mathbf{x}) \equiv \mathbb{E}(\mathrm{y}|\mathbf{x})$, then μ is a solution to

$$\min_{m \in \mathscr{M}} \mathbb{E}[(\mathrm{y} - m(\mathbf{x}))^2],$$

where $\mathscr{M}$ is the set of functions $m : \mathbb{R}^K \to \mathbb{R}$ such that $\mathbb{E}[m(\mathbf{x})^2] < \infty$. In other words, $\mu(\mathbf{x})$ is the best mean square predictor of y based on information contained in $\mathbf{x}$.

条件方差

以 x 为条件，y 的条件方差定义为：

$$\mathrm{Var}(\mathrm{y}|\mathbf{x}) \equiv \sigma^2(\mathbf{x}) \equiv \mathbb{E}[\{\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x})\}^2|\mathbf{x}] = \mathbb{E}(\mathrm{y}^2|\mathbf{x}) - [\mathbb{E}(\mathrm{y}|\mathbf{x})]^2.$$

根据条件方差的定义，容易证明：

$$\mathrm{Var}[a(\mathbf{x})\mathrm{y} + b(\mathbf{x})|\mathbf{x}] = [a(\mathbf{x})]^2\mathrm{Var}(\mathrm{y}|\mathbf{x}).$$

The ANOVA Theorem

$$\mathrm{Var}(y) = \mathbb{E}[\mathrm{Var}(y|\mathbf{x})] + \mathrm{Var}[\mathbb{E}(y|\mathbf{x})].$$

Proof.

The CEF decomposition property implies that

$$\mathrm{Var}(y) = \mathrm{Var}[\mathbb{E}(y|\mathbf{x})] + \mathrm{Var}(\varepsilon).$$

And, note that

$$\mathrm{Var}(\varepsilon) = \mathbb{E}(\varepsilon^2) = \mathbb{E}[\mathbb{E}(\varepsilon^2|\mathbf{x})] = \mathbb{E}[\mathrm{Var}(y|\mathbf{x})].$$

□

An Extension of the ANOVA Theorem

$$\mathrm{Var}(y|\mathbf{x}) = \mathbb{E}[\mathrm{Var}(y|\mathbf{x},\mathbf{z})|\mathbf{x}] + \mathrm{Var}[\mathbb{E}(y|\mathbf{x},\mathbf{z})|\mathbf{x}].$$

Proof.

Note that

$$\begin{aligned}
\mathrm{Var}(y|\mathbf{x}) &\equiv \mathbb{E}\left\{[y-\mathbb{E}(y|\mathbf{x})]^2|\mathbf{x}\right\} \\
&= \mathbb{E}\left\{[y-\mathbb{E}(y|\mathbf{x},\mathbf{z})+\mathbb{E}(y|\mathbf{x},\mathbf{z})-\mathbb{E}(y|\mathbf{x})]^2|\mathbf{x}\right\} \\
&= \mathbb{E}\left\{[y-\mathbb{E}(y|\mathbf{x},\mathbf{z})]^2|\mathbf{x}\right\}+\mathbb{E}\left\{[\mathbb{E}(y|\mathbf{x},\mathbf{z})-\mathbb{E}(y|\mathbf{x})]^2|\mathbf{x}\right\} \\
&\quad + 2\underbrace{\mathbb{E}\left\{[y-\mathbb{E}(y|\mathbf{x},\mathbf{z})]\cdot[\mathbb{E}(y|\mathbf{x},\mathbf{z})-\mathbb{E}(y|\mathbf{x})]|\mathbf{x}\right\}}_{=\,0}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathbb{E}\left\{[\mathbb{E}(y|\mathbf{x},\mathbf{z})-\mathbb{E}(y|\mathbf{x})]^2|\mathbf{x}\right\} &= \mathbb{E}\left\{[\mathbb{E}(y|\mathbf{x},\mathbf{z})-\mathbb{E}(\mathbb{E}(y|\mathbf{x},\mathbf{z})|\mathbf{x})]^2|\mathbf{x}\right\} \\
&= \mathrm{Var}[\mathbb{E}(y|\mathbf{x},\mathbf{z})|\mathbf{x}]. \\
\mathbb{E}\left\{[y-\mathbb{E}(y|\mathbf{x},\mathbf{z})]^2|\mathbf{x}\right\} &= \mathbb{E}\left\{\mathbb{E}\left\{[y-\mathbb{E}(y|\mathbf{x},\mathbf{z})]^2|\mathbf{x},\mathbf{z}\right\}|\mathbf{x}\right\} \\
&= \mathbb{E}[\mathrm{Var}(y|\mathbf{x},\mathbf{z})|\mathbf{x}].
\end{aligned}$$

□

From the Extension of the ANOVA Theorem, it is easy to know that

$$\mathbb{E}[\mathrm{Var}(\mathrm{y}|\mathbf{x})] \geq \mathbb{E}[\mathrm{Var}(\mathrm{y}|\mathbf{x}, \mathbf{z})]. \tag{1}$$

For any function $m(\cdot)$ define the mean squared error as

$$\mathrm{MSE}(\mathrm{y}; m) \equiv \mathbb{E}\left[(\mathrm{y} - m(\mathbf{x}))^2\right].$$

Then (1) can be loosely stated as $\mathrm{MSE}[\mathrm{y}; \mathbb{E}(\mathrm{y}|\mathbf{x})] \geq \mathrm{MSE}[\mathrm{y}; \mathbb{E}(\mathrm{y}|\mathbf{x}, \mathbf{z})]$. In other words, in the population one never does worse for predicting y when additional variables are conditioned on.

线性投影

- 由 CEF Prediction Property 定理可知 $\mu(\mathbf{x}) \equiv \mathbb{E}(\mathrm{y}|\mathbf{x})$ 是 best mean square predictor of y.
- $\mu(\mathbf{x})$ 一般是未知的，在实证分析中，通常采用线性函数近似 CEF:

$$\mu(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$$

Best Linear Predictor

The best linear predictor of y given $\mathbf{x}$ is

$$\mathbb{L}(\mathrm{y}|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$$

where $\boldsymbol{\beta}$ minimizes the mean squared prediction error

$$S(\boldsymbol{\beta}) = \mathbb{E}\left((\mathrm{y} - \mathbf{x}'\boldsymbol{\beta})^2\right).$$

The minimizer

$$\boldsymbol{\beta} = \underset{\boldsymbol{b}\in\mathbb{R}^K}{\arg\min}\ S(\boldsymbol{b})$$

is called the Linear Projection（线性投影）Coefficient.

定义 $e = \mathrm{y} - \mathbf{x}'\boldsymbol{\beta}$，由最优化的一阶条件，我们有：

$$\mathbb{E}(\mathbf{x}e) = \mathbf{0}.$$

于是，$\boldsymbol{\beta} = (\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1}\mathbb{E}(\mathbf{x}\mathrm{y})$.

- 如果 $\mathbf{x} = (1, x_1)'$，则 $\beta_1 = \dfrac{\mathrm{Cov}(\mathrm{y}, x_1)}{\mathrm{Var}(x_1)}, \alpha = \mathbb{E}(\mathrm{y}) - \beta_1\mathbb{E}(x_1)$.
- 如果 $\mathbf{x} = (1, x_1, \ldots, x_K)'$，其中 $K > 1$，则

$$\beta_k = \frac{\mathrm{Cov}(\mathrm{y}, \tilde{x}_k)}{\mathrm{Var}(\tilde{x}_k)}, \tag{2}$$

  其中，$\tilde{x}_k$ 是 x_k 对所有其他协变量投影得到的残差。
- It shows us that each coefficient in a multivariate regression is the bivariate slope coefficient for the corresponding regressor after partialing out all the other covariates.

To verify (2), 将

$$y = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k + \cdots + \beta_K x_K + e$$

代入 $\mathrm{Cov}(y, \tilde{x}_k)$，注意到：

- $\mathrm{Cov}(\tilde{x}_k, e) = 0$,
- $\mathrm{Cov}(\tilde{x}_k, x_j) = 0, \ j \neq k$,
- $\mathrm{Cov}(\tilde{x}_k, x_k) = \mathrm{Cov}(\tilde{x}_k, \tilde{x}_k + \mathbf{x}'_{-k}\boldsymbol{\beta}_{-k}) = \mathrm{Var}(\tilde{x}_k)$.[1]

于是,

$$\frac{\mathrm{Cov}(y, \tilde{x}_k)}{\mathrm{Var}(\tilde{x}_k)} = \frac{\beta_k \mathrm{Var}(\tilde{x}_k)}{\mathrm{Var}(\tilde{x}_k)} = \beta_k.$$

[1] $\mathbf{x}_{-k}$ 表示除 x_k 之外的所有协变量构成的向量。

Below we discuss three reasons why the vector of Linear Projection Coefficients might be of interest.

The Linear CEF Theorem (Regression Justification I)

Suppose the CEF is linear. Then $\mathbb{L}(\mathrm{y}|\mathbf{x})$ is it.

Proof.

假设 $\mathbb{E}(\mathrm{y}|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}^*$，其中，$\boldsymbol{\beta}^*$ 是 $K \times 1$ 的系数向量。由 CEF 分解定理,

$$\mathbb{E}[\mathbf{x}(\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x}))] = 0.$$

容易证明：$\boldsymbol{\beta}^* = \boldsymbol{\beta}$. $\square$

The Best Linear Predictor Theorem (Regression Justification II)

The function $\mathbf{x}'\boldsymbol{\beta}$ is the best linear predictor of y given $\mathbf{x}$ in an MSE sense.

The Regression CEF Theorem (Regression Justification III)

The function $\mathbf{x}'\boldsymbol{\beta}$ provides the minimum MSE linear approximation to $\mathbb{E}(\mathrm{y}|\mathbf{x})$, that is,

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{b}} \mathbb{E}\left\{(\mathbb{E}(\mathrm{y}|\mathbf{x}) - \mathbf{x}'\boldsymbol{b})^2\right\}.$$

Proof.

Write

$$\begin{aligned}(\mathrm{y} - \mathbf{x}'\boldsymbol{b})^2 &= \left\{(\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x})) + (\mathbb{E}(\mathrm{y}|\mathbf{x}) - \mathbf{x}'\boldsymbol{b})\right\}^2 \\ &= [\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x})]^2 + [\mathbb{E}(\mathrm{y}|\mathbf{x}) - \mathbf{x}'\boldsymbol{b}]^2 + 2[\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x})] \cdot [\mathbb{E}(\mathrm{y}|\mathbf{x}) - \mathbf{x}'\boldsymbol{b}].\end{aligned}$$

Therefore, minimize $\mathbb{E}\left\{(\mathbb{E}(\mathrm{y}|\mathbf{x}) - \mathbf{x}'\boldsymbol{b})^2\right\}$ is equivalent to minimize $\mathbb{E}\left\{(\mathrm{y} - \mathbf{x}'\boldsymbol{b})^2\right\}$. □

References

Joshua D. Angrist & Jörn-Steffen Pischke (2009)

Mostly Harmless Econometrics: An Empiricsist's Companion, Chapter 3.

Princeton University Press, 2009.

Bruce E. Hansen. (2020)

Econometrics, Section 2.18.

https://www.ssc.wisc.edu/ bhansen/econometrics/ Link